Stat and Data Analysis NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CW 4.1 – Correlation DATE: \_\_\_\_\_\_\_\_\_\_\_ BLOCK: \_\_\_\_

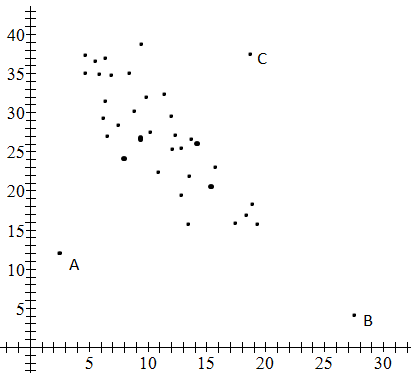
1. After conducting a survey at a pet store to see what impact having a pet had on the condition of the yard, a news reporter stated “There appears to be a strong correlation between the owning a pet(yes or no) and the condition of the yard(good or bad).” Comment on this observation- what is the problem with it?

2. A study by a prominent psychologist found a moderately strong positive association between the number of hours of sleep a person gets and the person’s ability to retain information.

1. Explain in the context of this problem what “positive association” means.
2. Hoping to improve academic performance, the psychologist recommended the school board allow students to take a nap prior to any assessment, because then they will definitely get higher scores. Discuss the psychologist’s recommendations- what is the problem with it?

3. Given the scatterplot on the right.

1. Describe the plot.



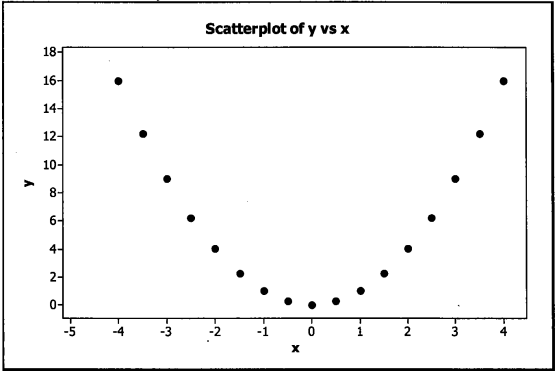
1. ***Ignoring the 3 marked outliers***, estimate the correlation. r = \_\_\_\_\_\_\_
2. Three points were labeled on the graph. For each one: if the point were removed, would the correlation become stronger, weaker, or remain about the same?

A: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ B: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ C: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Identify what is ***wrong*** with each of the following statements:

1. The correlation between Olympic gold medal times for the 800m hurdles and year is –0.66 seconds per year.
2. The correlation between Olympic gold medal times for the 100m dash and year is –1.37.
3. Since the correlation between Olympic gold medal times for the 800m hurdles and 100m dash is –0. 41, the correlation between times for the 100m dash and the 800m hurdles is +0.41.
4. If we were to measure Olympic gold medal times for the 800m hurdles in minutes instead of seconds, the correlation would be –0.66/60 = –0.011.

5. The following scatterplot shows a relationship between x and y that results in a correlation of r =0.

1. Describe the plot
2. Explain why r = 0 in this situation even though there is a ***strong*** relationship between the x and y variables.

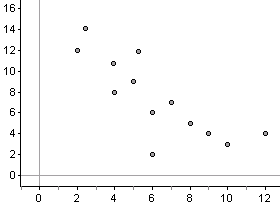
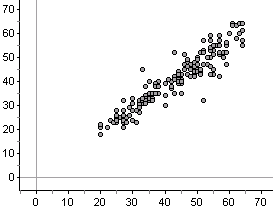
6. Correlation measures the strength of ***any*** association between two quantitative variables.

A. True B. False

7. Outliers in a scatterplot will:

1. Increase the correlation
2. Decrease the correlation
3. Change the sign of the correlation
4. All of the above depending on where the outlier is placed.

Given the following scatterplot which of the following would be the correct correlation:

8. 9.

A. -0.984 A. -0.921

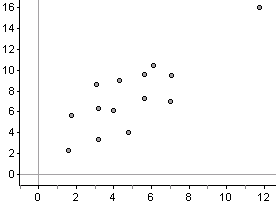
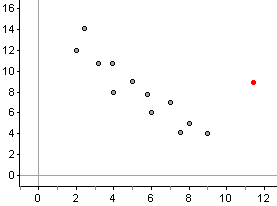
B. 0.034 B. -0.787

C. 0.678 C. 0.002

D. 0.939 D. 0.765

10. In the following scatterplot, there is an outlier. 11. In the following scatterplot, there is an outlier.

If the outlier were removed… If the outlier were removed…



1. The correlation would strengthen a) The correlation would strengthen
2. The correlation would weaken b) The correlation would weaken
3. The correlation would stay the same c) The correlation would stay the same
4. The correlation would change signs d) The correlation would change signs